

# Bayesian latent time joint mixed effect models for multicohort longitudinal data

Dan Li<sup>a</sup>, Samuel Iddi<sup>a,b</sup>, Wesley K. Thompson<sup>c</sup>, Michael C. Donohue<sup>a</sup>, for the Alzheimer's Disease Neuroimaging Initiative<sup>1</sup>

<sup>a</sup>*Alzheimer's Therapeutic Research Institute, University of Southern California, San Diego, USA*

<sup>b</sup>*Department of Statistics, University of Ghana, Legon-Accra, Ghana*

<sup>c</sup>*Department of Psychiatry, University of California, San Diego, USA*

---

## Abstract

Characterization of long-term disease dynamics, from disease-free to end-stage, is integral to understanding the course of neurodegenerative diseases such as Parkinson's and Alzheimer's; and ultimately, how best to intervene. Natural history studies typically recruit multiple cohorts at different stages of disease and follow them longitudinally for a relatively short period of time. We propose a latent time joint mixed effects model to characterize long-term disease dynamics using this short-term data. Markov chain Monte Carlo methods are proposed for estimation, model selection, and inference. We apply the model to detailed simulation studies and data from the Alzheimer's Disease Neuroimaging Initiative.

*Keywords:* hierarchical Bayesian models, joint mixed effects models, latent time shift, multicohort longitudinal data

---

## 1. Introduction

Disentangling the effects of normal aging from the effects of pathophysiology on neurobiological trajectories is crucial for predicting who will develop Alzheimer's or other neurodegenerative diseases. To make matters more difficult, the determination of neurobiological trajectories is typically based on data obtained over a relatively short time frame relative to the time span of normal aging and neurodegeneration [1]. There are four key barriers to the accurate characterization of longitudinal trajectories of neurodegeneration and pathology in older adults: 1) late-life trajectories can span 40 or more years; 2) there is substantial variation in the timing and temporal dynamics of late-life trajectories; 3) study selection criteria and sampling have a substantially different impact across age groups; and 4) there are potentially significant levels of censoring due to death and disability in late-life that cannot be treated as independent of pathology trajectories in most cases. Accounting for these issues

---

*Email address:* mdonohue@usc.edu (Michael C. Donohue)

<sup>1</sup>Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

is crucial to studying the onset, course and fluctuations of neurodegenerative diseases [2, 3], and for guiding when and how to target interventions and preventative approaches [4]. A better understanding of how and when to intervene necessitates the determination of dynamic networks of trajectories of markers of cognition, function, and pathology. Long-term disease dynamics are of great interest and importance, and have been hypothesized without rigorous methods.

We propose a latent time joint mixed effects model (LTJMM) for characterizing biomarker trajectories in aging. The model extends joint mixed effects models [5] to include an individual-specific latent time shift. Each individual’s latent time shift is shared across all of their outcomes and represents the extent of their long-term disease progression. The model is similar to others proposed for this application (e.g. Jedynak et al.[6] and Donohue et al.[7]). However, the proposed model can also accommodate covariates for fixed-effects and the Bayesian implementation allows flexible but rigorous interrogation of the posterior distribution to make inferences about long-term disease dynamics and the potential propagation of treatment effects from early biomarkers to downstream cognitive and functional measures. Estimation and inference are accomplished using Hamiltonian Markov Chain Monte Carlo as implemented in **Stan** [8] and the R package **rstan** (Stan Development Team[9], version 2.10.1). The **Stan** code model specifications and related tools for estimation and prediction are available as an additional R package from <https://bitbucket.org/mdonohue/ljtjmm>.

## 2. Methods

### 2.1. Latent time mixed-effects models for joint longitudinal responses

Suppose that  $p$  response variables are measured for  $n$  individuals at  $q$  follow-up times. Follow-up may differ from subject to subject and outcome to outcome, but we fix  $q$  to simplify notation. The measured outcomes for individual  $i$  at time  $j$  are denoted as  $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijp})'$ , which could be mixtures of binary, count, ordinal, and continuous outcomes. Assuming  $y_{ijk}$  has a distribution in the exponential family with canonical parameter  $\theta_{ijk}$ , the latent time joint mixed effects model (LTJMM) is given by

$$h_k(\theta_{ijk}) = \eta_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta}_k + \gamma_k(t_{ij} + \delta_i) + \alpha_{0ik} + \alpha_{1ik}t_{ij} + \varepsilon_{ijk}, \quad (1)$$

where  $h_k(\cdot)$  is a monotonic differentiable link function specific to the  $k$ -th outcome,  $\eta_{ijk}$  is the linear predictor for the  $k$ -th outcome from individual  $i$  at time  $j$ . The vector  $\mathbf{x}_{ijk}$  represents possibly time-varying covariates, and the vector  $\boldsymbol{\beta}_k$  is their corresponding regression coefficients. The  $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_k^2)$  is a measurement error term, which accounts for outcome-specific variance.

The parameters  $\alpha_{0ik}$  and  $\alpha_{1ik}$  are the subject and outcome specific random intercept and slope, and  $\delta_i$  is the subject-specific time shift shared among outcomes. The parameter  $\gamma_k$  corresponds to the outcome-specific slope with respect to “shifted” or “long-term” time. The time shift  $\delta_i$  quantifies the progression of the  $i$ -th individual relative to the population, and  $\alpha_{1ik}$  provides the information on whether individual  $i$  is evolving faster or slower than the average individual for outcome  $k$ .

In application of the model, we typically include a fixed effect for age, as a time-varying covariate, for each outcome. This admits two relevant rate-of-change parameters: one for biological age and one for “disease time.” This reflects the fact that individuals can reach the

same stage of disease at different ages, and allows independent effects of healthy aging and disease progression. The random effects accomodate additional subject to subject variation in the level and rate of change of disease markers. The link function,  $h_k$ , can be used to accomodate any distribution from the exponential family.

Some constraints must be placed on (1) to ensure model identifiability based on the observed data without relying on informative prior distributions. In particular, to ensure identifiability of  $\delta$ , the following two constraints are sufficient:

**Constraint 1**  $\sum_{k=1}^p \alpha_{0ik} = \sum_{k=1}^p \alpha_{1ik} = 0$  for all individuals  $i = 1, \dots, n$ ; and

**Constraint 2**  $\gamma_k > 0$  for all outcomes  $k = 1, \dots, p$ .

A proof of identifiability is available in the appendix. In practice, we find that it was sufficient to merely constrain  $\sum_{k=1}^p \alpha_{0ik} = 0$ , however the proof of identifiability requires that  $\alpha_{1ik}$  terms are also constrained. Identifiability also requires there are more subjects than outcomes ( $n > p$ ). Shape invariant models and self modeling regression (e.g. Kneip and Gasser[10]) similarly require positive first derivatives (Constraint 2). Model (1) can be modified to address more flexible nonlinear relationships over time by incorporating higher-order terms or splines [11], provided these extensions maintain monotonicity (Constraint 2).

Finally, we explore two different distribution assumptions for random effects, namely:

**1. Univariate Gaussian random effects**  $\alpha_{0ik} \sim \mathcal{N}(0, \sigma_{0k}^2)$  and  $\alpha_{1ik} \sim \mathcal{N}(0, \sigma_{1k}^2)$ ;

**2. Multivariate Gaussian random effects**  $\alpha_i \sim \mathcal{N}(\mathbf{0}, \Sigma)$ ,

where  $\Sigma$  is a covariance matrix of dimension  $2p$ . The latter assumption allows more direct exploration and inference regarding the correlation of response variables.

## 2.2. Prior specification

Since an improper prior may result in an improper posterior, we will use proper but weakly informative priors on all the model parameters [12]. The regression parameters  $\beta$  and  $\gamma$  are assigned independent weakly informative normal  $\mathcal{N}(0, 100)$  priors (truncated below by 0 for  $\gamma$ ). Without specific information, the half-Cauchy prior is a good default choice for scale parameters [12]. The standard deviations, i.e.,  $\sigma_\delta, \sigma_{01}, \dots, \sigma_{0p}, \sigma_{11}, \dots, \sigma_{1p}$  and  $\sigma$ , are given weakly informative half-Cauchy priors with a small scale, i.e., half-Cauchy  $(0, 2.5)$ . The prior variance 100 for the regression parameters is chosen to be sufficiently high to be vague enough, but sufficiently low to avoid slow mixing due to near impropriety of the posterior [13].

To ensure efficiency and arithmetic stability, we applied the Cholesky decomposition to the covariance matrix for the random effects,  $\Sigma$ , allowing us to model the standard deviations and correlations independently. We first decomposed the covariance matrix as  $\Sigma = \Lambda \Omega \Lambda$ , where  $\Lambda$  is a diagonal matrix with diagonal elements  $\sigma_{01}, \dots, \sigma_{0p}, \sigma_{11}, \dots, \sigma_{1p}$ , and  $\Omega$  is the correlation matrix with 1's on the diagonal and off-diagonal elements  $\rho$ . A Choleksy decomposition gives  $\Omega = \mathbf{L} \mathbf{L}'$ , where  $\mathbf{L}$  is a lower triangular matrix. Furthermore,  $\Sigma = \Lambda \mathbf{L} \mathbf{L}' \Lambda$ . During sampling, a draw is obtained from a multivariate Gaussian  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ , and then the random effects are calculated as  $\Lambda \mathbf{L} \mathbf{z}$ . Here,  $\mathbf{z}$  is regarded as a random reparameterization and independent of  $\rho$ . As suggested in Stan [14] manual, we imposed an LKJ prior on the correlation matrix [15].

### 2.3. Model comparison criteria

We use two model comparison criteria, namely, the widely applicable information criterion (WAIC) [16] and the leave-one-out cross-validation information criterion (LOOIC) [17, 18]. WAIC and LOOIC can be computed using the log-likelihood evaluated at the posterior simulations of the parameter values. Both have various advantages over the deviance information criterion (DIC). WAIC can be viewed as an improvement on DIC for Bayesian methods [19]. We refer to Gelman et al.[20] for the detailed reasons we prefer WAIC to DIC.

Approximate LOO can be computed using raw importance sampling (IS, Gelfand et al.[17]). Vehtari et al.[19] improved the LOO estimate using Pareto smoothed importance sampling (PSIS) method which provides a more accurate and reliable estimate than IS. The R package `loo` [21] provides tools for efficient computation of WAIC and LOOIC. The minimum WAIC and LOOIC indicate the best fit.

## 3. Results

### 3.1. Simulation study

We conducted a simulation study to assess the model performance and assess small sample bias. Data were generated from two candidate LTJMMs with identity link: (M1) Univariate Gaussian random effects; (M2) Multivariate Gaussian random effects. We simulated  $n = 400$  individuals with  $p = 4$  outcomes and  $q = 4$  time points each. The observation times for each individual were sampled from a Uniform distribution  $t_{ij} \sim \text{Uniform}(0, 10)$ . Additional simulation parameters were set to  $\beta = (1, 0.5, 2, 0.8)'$ ,  $\gamma = (0.2, 0.1, 0.25, 0.5)'$ ,  $\sigma_\delta = 4$ ,  $\sigma = (0.1, 0.2, 0.3, 0.25)'$ ,  $\sigma_{\alpha_0} = (0.5, 1, 0.8)'$ , and  $\sigma_{\alpha_1} = (1, 2, 1.5, 1)'$ .

Let  $\hat{\theta}_m$  denote the posterior estimate of a parameter  $\theta$  for the  $m$ -th simulation and let  $M$  denote the total number of simulations. We consider the following quantities to assess model performance: (i) total bias,  $\text{Bias} = \sum_{m=1}^M (\hat{\theta}_m - \theta_m) / M$ , (ii) total mean squared error of prediction,  $\text{MSPE} = \sum_{m=1}^M (\hat{\theta}_m - \theta_m)^2 / M$ , and (iii) coverage rate of the 95% credible intervals,  $C_{95}$ . Results are reported and discussed in Section 3.2. Each model was simulated  $M = 100$  times and both models were fit to each simulated data set.

For each model fit, two parallel Markov chains were run with dispersed initial values to diagnosis convergence. For each single chain, we ran 2 000 iterations and discarded the first 1 000 iterations as a warm-up phase, yielding a total of 2 000 samples for posterior analysis. The remaining samples were used to calculate posterior summaries of the parameters of interest and model comparison measurements. The potential scale reduction statistic  $\hat{R}$  [22], calculated by Stan was used to verify posterior convergence.

### 3.2. Simulation study results

The estimated potential scale reduction factors were below 1.1 for all parameters, indicating successful convergence. Table 1 summarizes the measures of total bias, MSPE,  $C_{95}$ , and the percentage of best performance in terms of WAIC and LOOIC for each scenario. In Table 1, we can see the advantage of M2 under two different scenarios with better coverage rates of the 95% credible intervals for most parameters and larger percentage of best performance. It may not be surprising that M2 is preferred even when M1 is the true model, since M2 accommodates the association between outcomes and includes M1 as a special case. For

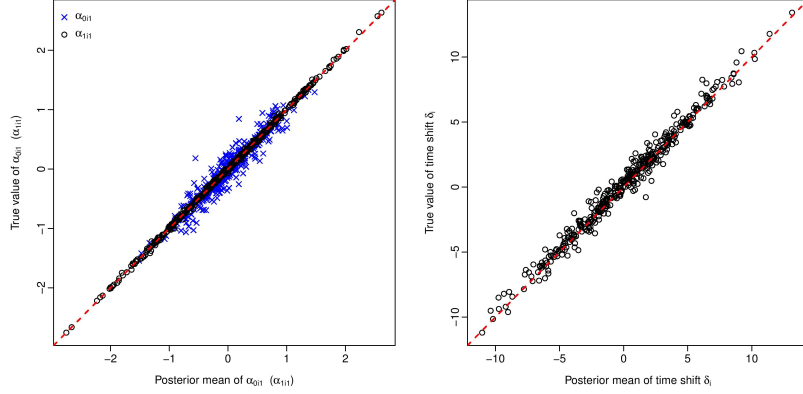


Figure 1: Results of one simulated data example with 4 outcomes, 400 individuals, and at 4 time points. (a) True value versus posterior mean of random intercepts and slopes for the first outcome. Black circle and blue cross indicate the slope and intercept, respectively. (b) True value and posterior mean of latent time shifts, and these are indicated in black circles. The red dashed line is an identity line (points lie close to the line indicating a good fit).

both scenarios, the regression parameters were estimated with low bias and MSPE, indicating good performance in terms of prediction error. Figure 1 plots the true versus estimated individual intercepts and slopes for the first outcome, and the true versus estimated time shifts of one simulated data set. Both plots demonstrate agreement between the true and estimated values.

### 3.3. Application to Alzheimer’s Disease Neuroimaging Initiative

The LTJMM was fit to data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI), which has followed volunteers diagnosed as cognitively healthy or with varying degrees of cognitive impairment since 2005 [23]. The ADNI battery includes serial neuroimaging, cerebrospinal fluid (CSF), and other biomarkers; and clinical and neuropsychological assessments. Participants returned for repeated assessments at six months, one year, and every year thereafter. Seven outcome measures were included in the model: CSF tau and amyloid beta ( $A\beta$ ) 1-42; PET imaging of amyloid deposition and glucose metabolism in the brain; volumetric magnetic resonance imaging (vMRI) of the hippocampus; the 13 item Alzheimer’s Disease Assessment Scale (ADAS13); and the Functional Activities Questionnaire (FAQ). Fixed effect covariates for each outcome included age, carriage of the  $APOE\epsilon 4$  allele, sex, and education. The model was fit with and without assuming a multivariate Gaussian distribution for the random effects. Three parallel Markov chains were run for 8 000 iterations and the first 4 000 warm-up iterations were discarded. Every fourth value of the remaining part of each chain were stored to reduce level of correlation, yielding a total of 3 000 samples for posterior analysis. The posterior mean and the 95% credible intervals were calculated using the obtained samples for each parameter.

One goal of the analysis is to compare the long-term trends of the outcomes on a comparable scale, and make conclusions about the temporal ordering of their emergence. With this goal in mind, the outcome measures were transformed first to quantiles, and the quantiles were then transformed by the inverse Gaussian quantile function. All the transformed outcomes were then modeled as Gaussian with identity link. ADNI subjects were diagnosed at their first visit with normal cognition (24%), subjective memory concern (6%), early

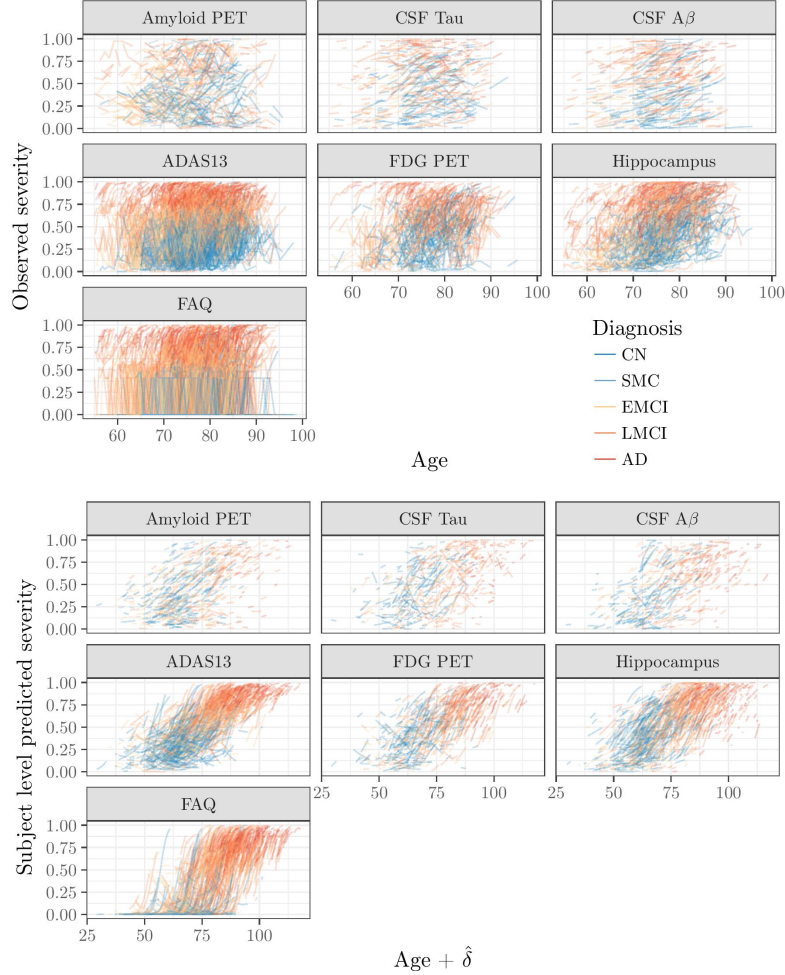


Figure 2: Subject-level observed and predicted severity. The top panel shows spaghetti plots of the observed quantiles of each outcome from all subjects in the Alzheimer’s Disease Neuroimaging Initiative with respect to their age over time. The bottom panels shows modeled trajectories for these same subjects from the fitted LTJMM with respect to the sum of age and estimated latent time,  $\delta$ . The colors indicate diagnostic severity at first observation, from cognitively normal (blue) through dementia (red).

Abbreviations: ADAS13, Alzheimer’s Disease Assessment Scale (13 Item version); FDG, fluorodeoxyglucose; PET, positron emission tomography; CSF, cerebrospinal fluid; FAQ, Functional Activities Questionnaire; CN, cognitively normal; SMC, subjective memory concern; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, probable Alzheimer’s Disease with mild to moderate dementia.

Table 1: Simulation study results. One hundred simulated data sets were generated under model M1 (univariate Gaussian random effects) and model M2 (multivariate Gaussian random effects). Both models were fit to all 200 data sets. The best performance was determined by the lowest WAIC and LOOIC among the two models for each data and summarized as “% best” over the 100 simulations.

Parameter	True Model: M1						True Model: M2					
	Fitted Model: M1 Bias MSPE $C_{95}$	Fitted Model: M1 Bias MSPE $C_{95}$	Fitted Model: M1 Bias MSPE $C_{95}$	Fitted Model: M2 Bias MSPE $C_{95}$	Fitted Model: M2 Bias MSPE $C_{95}$	Fitted Model: M2 Bias MSPE $C_{95}$	Fitted Model: M1 Bias MSPE $C_{95}$	Fitted Model: M1 Bias MSPE $C_{95}$	Fitted Model: M1 Bias MSPE $C_{95}$	Fitted Model: M2 Bias MSPE $C_{95}$	Fitted Model: M2 Bias MSPE $C_{95}$	Fitted Model: M2 Bias MSPE $C_{95}$
$\beta_1$	-0.0008	0.0021	0.95	-0.0005	0.0021	0.96	< .0001	0.0023	0.93	0.0016	0.0024	0.95
$\beta_2$	0.0032	0.0032	0.88	0.0025	0.0033	0.95	0.0037	0.0032	0.92	0.0024	0.0032	0.96
$\beta_3$	0.0057	0.0057	0.90	0.0054	0.0056	0.92	0.0062	0.0056	0.92	0.0073	0.0057	0.90
$\beta_4$	0.0042	0.0139	0.94	0.0047	0.0149	0.94	0.0053	0.0180	0.90	0.0089	0.0181	0.95
$\gamma_1$	0.0035	0.0004	0.91	0.0066	0.0005	0.89	0.0020	0.0007	0.81	0.0115	0.0008	0.91
$\gamma_2$	0.0028	0.0003	0.88	0.0062	0.0003	0.92	0.0024	0.0004	0.90	0.0110	0.0006	0.90
$\gamma_3$	0.0056	0.0009	0.86	0.0098	0.0009	0.91	0.0063	0.0014	0.77	0.0168	0.0016	0.92
$\gamma_4$	0.0089	0.0024	0.92	0.0091	0.0022	0.93	0.0068	0.0034	0.83	0.0139	0.0026	0.91
$\sigma_1$	0.0005	< .0001	0.95	0.0022	< .0001	0.95	0.0004	< .0001	0.95	0.0031	0.0002	0.96
$\sigma_2$	0.0006	< .0001	0.95	0.0018	0.0002	0.95	-0.0006	< .0001	0.95	0.0001	< .0001	0.94
$\sigma_3$	0.0036	< .0001	0.95	0.0014	< .0001	0.96	0.0008	< .0001	0.96	0.0041	0.0002	0.96
$\sigma_4$	0.0033	< .0001	0.94	0.0029	0.0004	0.97	0.0044	< .0001	0.92	0.0022	0.0001	0.98
$\sigma_\delta$	-0.0385	0.1516	0.90	-0.0783	0.1447	0.89	-0.0010	0.2268	0.86	-0.1307	0.2325	0.89
$\sigma_{01}$	-0.0001	0.0004	0.96	-0.0013	0.0007	0.97	0.0092	0.0005	0.93	-0.0037	0.0017	0.94
$\sigma_{02}$	-0.0143	0.0016	0.90	0.0126	0.0020	0.96	0.0052	0.0017	0.94	0.0117	0.0017	0.93
$\sigma_{03}$	-0.0313	0.0018	0.90	0.0045	0.0009	0.99	-0.0113	0.0011	0.94	-0.0007	0.0011	0.97
$\sigma_{11}$	0.0044	0.0010	0.97	0.0118	0.0012	0.96	0.0028	0.0011	0.97	0.0038	0.0012	0.96
$\sigma_{12}$	0.0055	0.0046	0.97	0.0183	0.0051	0.97	0.0023	0.0042	0.94	-0.0038	0.0041	0.95
$\sigma_{13}$	0.0016	0.0029	0.94	0.0125	0.0032	0.94	0.0011	0.0020	0.98	-0.0075	0.0020	0.98
$\sigma_{14}$	-0.0005	0.0012	0.95	0.0077	0.0015	0.94	-0.0017	0.0012	0.95	-0.0028	0.0012	0.97
% best	9%			91%			8%			92%		

mild cognitive impairment (17%), late mild cognitive impairment (32%), and mild to moderate dementia (19%). A weighted quantile transformation [24] was used so that quantiles approximate the quantiles from a sample with equal numbers of each diagnosis.

We fit both models described in Section 2.1 to examine gains by incorporating correlated random effects. Table 2 presents model comparison results, and posterior estimates of the key model parameters for the best model. As shown in Table 2(b), the model assuming multivariate Gaussian random effects was preferred, with lower DIC, WAIC and LOOIC.

Figure 2 shows the subject-level observations (top) and predictions (bottom) according to age. It is clear from the the observations, that age explains only a small proportion of the variance in these outcomes. The bottom panel shows that the predictions provide a reasonable smooth of the observations, and that latent time provides a reasonable ordering of individuals according to disease severity. The posterior mean (95% credible interval) for the latent time parameter was 14.2 (12.7 to 16.1) years (Table 2). Figure 3 shows a density plot for the posterior mean of the subject-specific latent time by diagnosis at first ADNI visit. Diagnosis, which is a somewhat subjective interpretation of the clinical presentation (excluding CSF and imaging data) of the individuals by their physician, was not included as a covariate in the model. Figure 3 shows that the latent time estimates are temporally sorting individuals in a manner that is consistent with physician diagnosis. Latent time estimates provide a continuous alternative to diagnosis which is objectively derived from a comprehensive model of longitudinal measures of disease.

Figure 4 shows the posterior mean of correlation parameters for random intercepts and slopes. Not surprisingly, we see strong correlations between change in cognitive tests

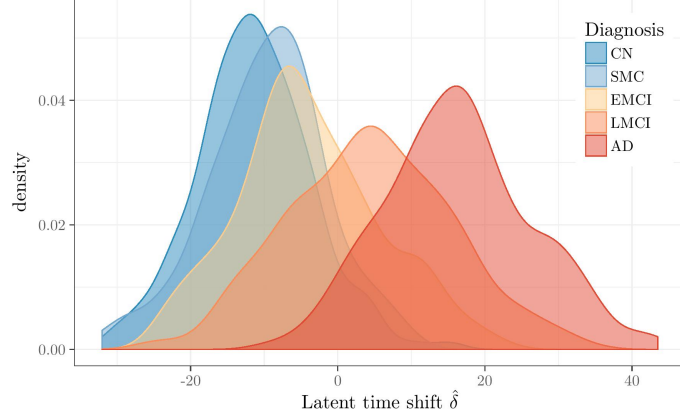


Figure 3: Distribution of the subject-specific latent time shifts. The estimated latent time shifts are colored by baseline diagnostic group, a variable not included in the model. This plot suggests that the time shifts are well aligned and consistent with diagnostic criteria. The density plot also demonstrates that there is much overlap of the diagnostic criteria with respect latent time.

Abbreviations: CN, cognitively normal; SMC, subjective memory concern; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, probable Alzheimer’s Disease with mild to moderate dementia.

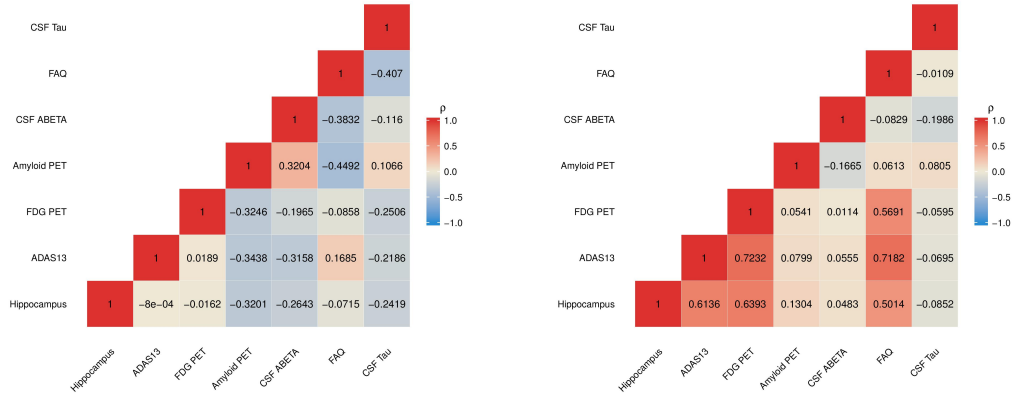


Figure 4: Posterior mean correlations among random intercepts (top) and slopes (bottom).

Abbreviations: ADAS13, Alzheimer’s Disease Assessment Scale (13 Item version); FDG, fluorodeoxyglucose; PET, positron emission tomography; CSF, cerebrospinal fluid; FAQ, Functional Activities Questionnaire; ADNI, Alzheimer’s Disease Neuroimaging Initiative.



(ADAS13) and the function (FAQ). However we also, see strong correlation between measures of symptomatic change (cognition and function) and biomarkers (hippocampal atrophy and glucose metabolism [FDG PET]). The two amyloid measures, CSF  $A\beta$  and amyloid PET, show a moderate positive correlation for random intercepts and weakly negative correlation of change.

Figure 5 displays the population-level predicted trajectories. The depicted curve are for female APOE $\epsilon$ 4 carriers with the ADNI mean education. Recall that age and latent time contribute to the model independently. For the sake of these predictions, age is calibrated so that the ADAS13 trajectory attains the ADNI mean ADAS13 at the ADNI mean age at the first visit. The bottom panel shows the same trajectories for progressive Alzheimer’s (red triangles) with contrasting trajectories for healthy aging (blue dots). To obtain estimates for healthy aging, the effect of latent time is forced to be zero to isolate the effect of age.

## 4. Discussion

We explored sampling the random effects from both univariate and multivariate Gaussian distributions, and found the multivariate to always be preferred by model selection criteria. The model with multivariate random effects has the advantage of providing additional parameters for the correlation among outcomes (random intercepts) and among change in outcomes (random slopes). However, fitting the model with univariate random effects was often computationally faster, so it may have advantages in practice. We plan to explore extension of the model which include monotone smooth functions of latent time, and other random effects options.

The LTJMM provides a parsimonious extension of the existing family of joint generalized linear mixed effects models with a subject-specific latent time parameter that accommodates the temporal heterogeneity common to multicohort longitudinal data in late-life neurodegenerative diseases. The model provides key insights and inference regarding the evolution of disease markers over a period of time that is longer than the period of observation. The framework allows consideration of the independent effects of healthy aging and disease progression.

These novel features can be leveraged to improve subject-level prediction and better understand long-term disease dynamics. In particular we plan to leverage the model to help improve clinical trial design. For example, given an estimate of the short-term effect of a treatment on a biomarker, the model can be interrogated to estimate the downstream effects on cognition and function. The model can also be used to help identify populations expected to experience the maximum benefit from a given intervention. The LTJMM provides a flexible framework to begin to explore these and many other applications.

## 5. Supplementary material

The **Stan** code model specifications and related code and functions for simulation, estimation and prediction are available as an R package from <https://bitbucket.com/mdonohue/ltjmm>. Interactive convergence plots of the model fit to Alzheimer’s data is available from [https://shiny.atrihub.org/public/adni\\_ltjmm/](https://shiny.atrihub.org/public/adni_ltjmm/).

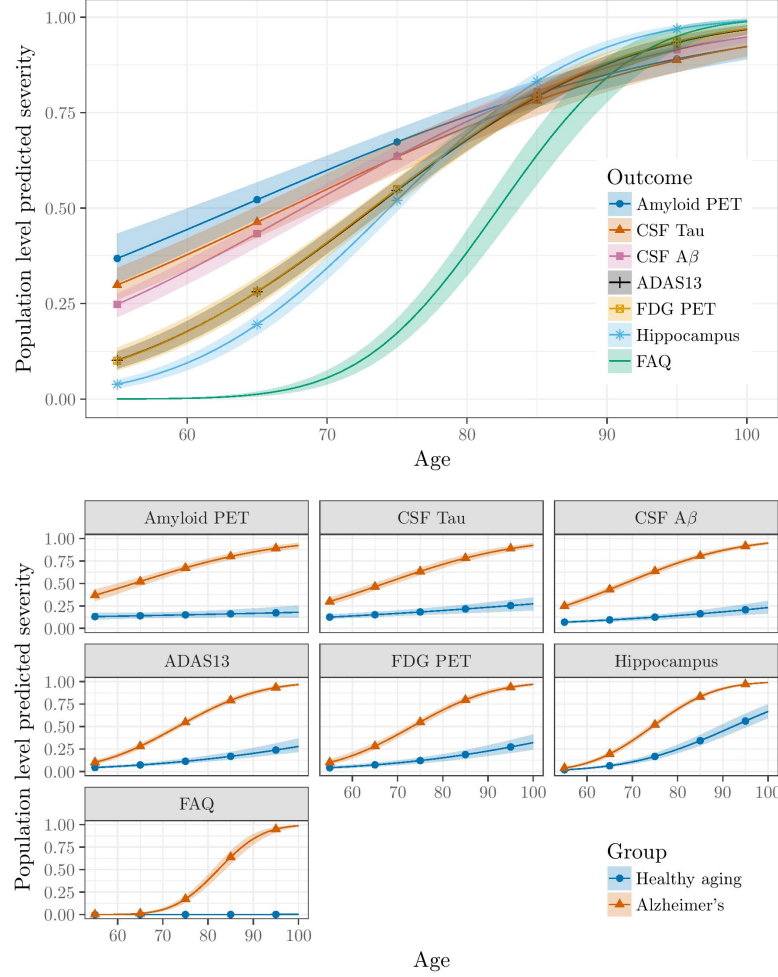


Figure 5: Modeled population-level severity. The top panel shows the modeled average trajectories. The depicted evolution is for female APOE $\epsilon$ 4 carriers with the ADNI mean education. Age is calibrated so that the ADAS13 trajectory attains the ADNI mean ADAS13 at the ADNI mean age at baseline. The bottom panel shows the same trajectories for progressive Alzheimers (red triangles) with contrasting trajectories for healthy aging (blue dots). For the latter, the effect of latent time is forced to be zero to isolate the effect of age.

Abbreviations: ADAS13, Alzheimer's Disease Assessment Scale (13 Item version); FDG, fluorodeoxyglucose; PET, positron emission tomography; CSF, cerebrospinal fluid; FAQ, Functional Activities Questionnaire; ADNI, Alzheimer's Disease Neuroimaging Initiative.

Table 2: Results of (a) posterior estimates of parameters, and (b) model selection for LTJMM fit to seven outcomes from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

(a) Posterior mean and 95% credible intervals of parameters for the proposed LTJMM.

Parameter	Posterior Mean	95% Credible Interval	Parameter	Posterior Mean	95% Credible Interval
<b>Hippocampal volume</b>			<b>CSF Amyloid Beta 1-42</b>		
Intercept	-4.102	(-4.516, -3.688)	Intercept	-1.454	(-2.061, -0.872)
Age	0.056	(0.052, 0.06)	Age	0.016	(0.011, 0.022)
APOE $\epsilon$ 4	0.317	(0.258, 0.394)	APOE $\epsilon$ 4	0.802	(0.704, 0.888)
Sex	-0.288	(-0.364, -0.215)	Sex	-0.028	(-0.116, 0.049)
Education	-0.003	(-0.015, 0.008)	Education	-0.012	(-0.03, 0.005)
Latent time	0.035	(0.03, 0.04)	Latent time	0.035	(0.03, 0.039)
$\sigma_1$	0.137	(0.134, 0.141)	$\sigma_5$	0.021	(0.019, 0.022)
<b>Alzheimer's Disease Assessment Scale</b>			<b>Functional Activities Questionnaire</b>		
Intercept	-1.089	(-1.57, -0.646)	Intercept	-3.012	(-3.934, -2.012)
Age	0.025	(0.02, 0.03)	Age	0.035	(0.025, 0.045)
APOE $\epsilon$ 4	0.434	(0.355, 0.514)	APOE $\epsilon$ 4	0.861	(0.678, 1.021)
Sex	-0.246	(-0.314, -0.168)	Sex	-0.335	(-0.471, -0.196)
Education	-0.051	(-0.064, -0.038)	Education	-0.061	(-0.089, -0.037)
Latent time	0.045	(0.039, 0.05)	Latent time	0.096	(0.085, 0.107)
$\sigma_2$	0.327	(0.32, 0.333)	$\sigma_6$	0.842	(0.826, 0.858)
<b>FDG PET</b>			<b>CSF Tau</b>		
Intercept	-1.820	(-2.348, -1.3)	Intercept	-0.766	(-1.265, -0.202)
Age	0.028	(0.023, 0.034)	Age	0.012	(0.007, 0.017)
APOE $\epsilon$ 4	0.453	(0.363, 0.54)	APOE $\epsilon$ 4	0.621	(0.505, 0.72)
Sex	-0.181	(-0.259, -0.103)	Sex	0.073	(-0.036, 0.19)
Education	-0.026	(-0.041, -0.01)	Education	-0.029	(-0.048, -0.007)
Latent time	0.042	(0.037, 0.048)	Latent time	0.032	(0.028, 0.036)
$\sigma_3$	0.294	(0.283, 0.304)	$\sigma_7$	0.016	(0.014, 0.017)
<b>Amyloid PET</b>			<b>Standard deviation of latent time</b>		
Intercept	-0.685	(-1.328, -0.013)	$\sigma_8$	14.262	(12.684, 16.105)
Age	0.004	(-0.002, 0.011)			
APOE $\epsilon$ 4	0.785	(0.669, 0.912)			
Sex	0.159	(0.067, 0.247)			
Education	-0.007	(-0.029, 0.014)			
Latent time	0.035	(0.03, 0.04)			
$\sigma_4$	0.289	(0.273, 0.307)			

(b) Model comparison results for the LTJMMs under two different assumptions of random effects.

Model	DIC	WAIC	LOOIC
1	-353.2055	-624.2033	-623.5437
2	-364.9088	-640.0899	-639.3916

## Acknowledgment

We are grateful to the ADNI study volunteers and their families. This work was supported by Biomarkers Across Neurodegenerative Disease (BAND-14-338179) grant from the Alzheimer’s Association, Michael J. Fox Foundation, and Weston Brain Institute; and National Institute on Aging grant R01-AG049750. Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bio-engineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## References

- [1] W. K. Thompson, J. Hallmayer, R. O’Hara, Alzheimer’s Disease Neuroimaging Initiative, Design considerations for characterizing psychiatric trajectories across the lifespan: application to effects of APOE- $\epsilon$ 4 on cerebral cortical thickness in Alzheimer’s disease, *The American Journal of Psychiatry* 168 (2011) 894–903.
- [2] C. J. Jack, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, J. Q. Trojanowski, Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade, *The Lancet. Neurology* 9 (2010) 119–128.
- [3] C. J. Jack, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, S. L. M., P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, J. Q. Trojanowski, Tracking pathophysiological processes in Alzheimer’s disease: an updated hypothetical model of dynamic biomarkers, *The Lancet. Neurology* 12 (2013) 207–216.
- [4] R. Sperling, S. Salloway, D. J. Brooks, D. Tampieri, J. Barakos, N. C. Fox, M. Raskind, M. Sabbagh, L. S. Honig, A. P. Porsteinsson, I. Lieberburg, H. M. Arrighi, K. A. Morris, Y. Lu, E. Liu, K. M. Gregg, R. H. Brashear, G. G. Kinney, R. Black, M. Grundman, Amyloid-related imaging abnormalities (ARIA) in Alzheimer’s disease patients treated with bapineuzumab: A retrospective analysis, *Lancet Neurology* 11 (2012) 241–249.
- [5] N. M. Laird, J. H. Ware, Random-effects models for longitudinal data, *Biometrics* 38 (1982) 963–974.
- [6] B. M. Jernak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jernak, B. Caffo, J. L. Prince, et al., A computational neurodegenerative disease progression score: Method and results with the alzheimer’s disease neuroimaging initiative cohort, *Neuroimage* 63 (3) (2012) 1478–1486.

- [7] M. C. Donohue, H. Jacqmin-Gadda, M. Le Goff, R. G. Thomas, R. Raman, A. C. Gamst, L. A. Beckett, C. R. Jack, Jr, M. W. Weiner, J. F. Dartigues, P. S. Aisen, Alzheimer's Disease Neuroimaging Initiative, Estimating long-term multivariate progression from short-term data, *Alzheimer's & Dementia : the Journal of the Alzheimer's Association* 10 (2014) S400–410.
- [8] B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, A. Riddell, Stan: a probabilistic programming language, *Journal of Statistical Software*(In press).
- [9] Stan Development Team, Rstan: the R interface to stan, version 2.10.1, <http://mc-stan.org> (2016).
- [10] A. Kneip, T. Gasser, Convergence and consistency results for self-modeling nonlinear regression, *The Annals of Statistics* 16 (1) (1988) 82–112.
- [11] D. Ruppert, M. P. Wand, C. J. Raymond, *Semiparametric regression*, Cambridge University Press, New York, 2003.
- [12] A. Gelman, Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis* 1 (2006) 515–534.
- [13] R. Natarajan, C. E. McCulloch, Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference?, *Journal of Computational and Graphical Statistics* 7 (1998) 267–277.
- [14] Stan Development Team, Stan modeling language users guide and reference manual, version 2.12.0 (2016).  
URL <http://mc-stan.org/>
- [15] D. Lewandowski, D. Kurowicka, H. Joe, Generating random correlation matrices based on vines and extended onion method, *Journal of Multivariate Analysis* 100 (2009) 1989–2001.
- [16] S. Watanabe, Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory, *Journal of Machine Learning Research* 11 (2010) 3571–3594.
- [17] A. E. Gelfand, D. K. Dey, H. Chang, Model determination using predictive distributions with implementation via sampling-based methods, in: J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), *Bayesian Statistics 4*, Oxford University Press, 1992, pp. 147–167.
- [18] A. E. Gelfand, Model determination using sampling-based methods, in: W. R. Gilks, S. Richardson, D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 1996, pp. 145–162.
- [19] A. Vehtari, A. Gelman, J. Gabry, Practical bayesian model evaluation using leave-one-out cross-validation and WAIC, *arXiv: 1507.04544v5* (2016).
- [20] A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for bayesian models, *Statistics and Computing* 24 (2014) 997–1016.
- [21] A. Vehtari, A. Gelman, J. Gabry, loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models, *r package version 0.1.6* (2016).  
URL <https://github.com/jgabry/loo>
- [22] A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences, *Statistical Science* 7 (1992) 457–472.
- [23] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, J. C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, M. W. Weiner, Alzheimer's disease neuroimaging initiative (adni): clinical characterization, *Neurology* 74 (3) (2010) 201.
- [24] F. E. Harrell Jr, Hmisc: Harrell Miscellaneous, *r package version 3.17-4* (2016).  
URL <https://CRAN.R-project.org/package=Hmisc>

## Appendix A. Proof of the Identifiability

Suppose that  $p$  response variables are measured for  $n$  individuals at  $q$  follow-up times. The measured outcomes for individual  $i$  at time  $j$  are denoted as  $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijp})'$ . The latent time joint mixed effects model is given by

$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta}_k + \gamma_k(t_{ij} + \delta_i) + \alpha_{0ik} + \alpha_{1ik}t_{ij} + \varepsilon_{ijk}, \quad (\text{A.1})$$

The model can be re-expressed as

$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta}_k + t_{ij}\gamma_k + \tilde{\alpha}_{0ik} + \tilde{\alpha}_{1ik}t_{ij} + \varepsilon_{ijk} \quad (\text{A.2})$$

where  $\tilde{\alpha}_{0ik} = \gamma_k\delta_i + \alpha_{0ik}$  and  $\tilde{\alpha}_{1ik} = \gamma_k + \alpha_{1ik}$ . Identifiability, up to  $\tilde{\alpha}_{0ik}$  and  $\tilde{\alpha}_{1ik}$  follows from the identifiability of standard linear mixed effects models (see Wang 2013<sup>2</sup> for a discussion) provided  $(\mathbf{x}', \mathbf{t})$  has full column rank where  $\mathbf{t}$  is the vector formed by  $t_{ij}$ ; and  $\mathbf{t}$  is linearly independent from the vector of 1s. It remains to show that:

$$\tilde{\alpha}_{0ik} = \gamma_k\delta_i + \alpha_{0ik} \quad (\text{A.3})$$

$$\tilde{\alpha}_{1ik} = \gamma_k + \alpha_{1ik} \quad (\text{A.4})$$

$$0 = \sum_{k=1}^p \alpha_{0ik} \quad (\text{A.5})$$

$$0 = \sum_{k=1}^p \alpha_{1ik} \quad (\text{A.6})$$

has a unique solution, where  $\tilde{\alpha}_{0ik}$  and  $\tilde{\alpha}_{1ik}$  are known; and  $\delta_i$ ,  $\gamma_k$ ,  $\alpha_{0ik}$ , and  $\alpha_{1ik}$  are unknown. There are  $2p + 1$  unknown subject-specific parameters and  $p$  additional  $\gamma_k$  parameters that are merely outcome-specific, for a total of  $n(2p + 1) + p$  unknown parameters. There are  $n(2p + 2)$  constraints. Therefore,  $\delta_i$ ,  $\gamma_k$ ,  $\alpha_{0ik}$ , and  $\alpha_{1ik}$  are uniquely determined provided

$$n(2p + 2) > n(2p + 1) + p, \text{ or} \quad (\text{A.7})$$

$$n > p. \quad (\text{A.8})$$

---

<sup>2</sup>Wang, W. (2013). Identifiability of linear mixed effects models. *Electronic Journal of Statistics*, 7, 244-263.